DOCUMENT RESUME

ED 304 463 TM 012 835

AUTHOR Lautenschlager, Gary J.
TITLE Determining the Number of Principal Components to Retain via Parallel Analysis: Alternatives to Monte Carlo Analyses.
PUB DATE Aug 88
NOTE 37p.; Paper presented at the Annual Meeting of the American Psychological Association (96th, Atlanta, GA, August 12-16, 1988).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Evaluative/Feasibility (142) -- Statistical Data (110)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Equations (Mathematics); *Estimation (Mathematics); Evaluation Criteria; *Monte Carlo Methods; Regression (Statistics); Statistical Analysis
IDENTIFIERS Eigenvalues; Interpolation; *Parallel Analysis (Horn); *Principal Components Analysis

ABSTRACT

The parallel analysis method for determining the number of components to retain in a principal components analysis has received a recent resurgence of support and interest. However, researchers and practitioners desiring to use this criterion have been hampered by the required Monte Carlo analyses needed to develop the criteria. Two recent attempts at presenting regression estimation methods to determine eigenvalues were found to be deficient in several respects, and less accurate in general, than a simple linear interpolation of tabled random data eigenvalues. Tables are presented which permit accurate and easy determination of the parallel analysis criteria within a range of sample sizes (N=50 through 1,000) and number of variables (P=5 through 50) covered by the tables. A total of 12,000 unique data sets was created. The generated data provided the empirical criteria for comparison of regression equation estimates using S. J. Allen and R. Hubbard's (1986) estimation equations with those revised by G. J. Lautenschlager et al. Twelve tables present average eigenvalues and values for interpolations. (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED304463

Determining the Number of Principal Components to Retain via Parallel Analysis: Alternatives to Monte Carlo Analyses

Gary J. Lautenschlager

University of Georgia

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

✓ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

GARY LAUTENSCHLAGER

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Address correspondence to:

Gary J. Lautenschlager

Department of Psychology

University of Georgia

Athens, GA 30602

(404) 542-3054

Paper presented at the 96th American Psychological Association Meetings, Atlanta, GA, August 12-16, 1988.

TM012835

RUNNING HEAD: PARALLEL ANALYSIS CRITERIA

Abstract

The parallel analysis method for determining the number of components to retain in a principal components analysis has received a recent resurgence of support and interest. However, researchers and practitioners desiring to use this criterion have been hampered by the required Monte Carlo analyses needed to develop the criteria. Two recent attempts at presenting regression estimation methods to determine eigenvalues were found to be deficient in several respects, and less accurate in general, than a simple linear interpolation of tabled random data eigenvalues. Tables are presented which permit accurate and easy determination of the parallel analysis criteria within a range of sample sizes (N = 50 through 1000) and number of variables (P = 5 through 50) covered by the tables.

Alternatives to Conducting Monte Carlo Analyses for Determining Parallel Analysis Criteria

Researchers using principal components analysis, either as a data reduction method or as a preliminary step in common factor analysis, have available a number of criteria for determining the number of factors to retain for rotation and/or interpretation. Perhaps the most prevalent method in use is Kaiser's (1960) criterion of retaining eigenvalues greater than one for determining the number of components, or common factors, to retain. However, applying Kaiser's rule, or any of several other rules that are available, may not be as accurate as using the parallel analysis (PA) criterion. It has been suggested that both the minimum average partial correlation method (Velicer, 1976) and the PA method might both be employed to reach a decision about the number of components to retain (Zwick and Velicer, 1986). Cliff (1988) has also called into question the use of eigenvalues-greater-than-one to determine the number of components or common factors. Horn (1965) initially proposed the PA method as a sample-based adaption of Kaiser's (1960) rule of retaining all eigenvalues greater than or equal to 1.00. The basic rationale underlying the PA criterion being that "meaningful" components extracted from actual sample data should tend to have eigenvalues larger in size than eigenvalues of the same order obtained from random normal variates generated to simulate the same sample size and number of variables. Since it is the expected value of eigenvalues of random data that are used for the comparison, it is necessary to simulate many such datasets and average the respective eigenvalues obtained to estimate the expected values. A stumbling block for those desiring to implement the PA criterion in practice is the required number of Monte Carlo analyses.

Regression Estimates of Eigenvalues

Allen and Hubbard (1986) have presented a means by which parallel analysis could be made more accessible to practitioners. Their work was based on earlier work by Montanelli and Humphreys (1976; Humphreys & Montanelli, 1975) which had presented a general form of a regression equation to be used for estimating the eigenvalues of random data correlation matrices with squared multiple correlations on the diagonal. Allen and Hubbard (1986) derived a general form of a regression equation designed to predict values of all but the two smallest (last) eigenvalues of a random data correlation matrix with unities on the diagonal. The Allen and Hubbard equation is:

$$[1] \quad \log(\lambda_k) = a_k + b_k \log(N-1) + c_k \log\{(P-k-1)(P-k+2)/2\} + d_k \log(\lambda_{k-1})$$

where log is the natural logarithm, N is sample size, P is the number of variables, k [$k = 1,2,\ldots,(P-2)$] indexes the k_{th} eigenvalue λ_k, ($\lambda_0 = 1.0$) and a_k, b_k, c_k and d_k are regression model parameters. The term weighted by c_k was derived from a degrees-of-freedom rationale offered by Bartlett (1951) and Lawley (1956). The inclusion of this term necessarily restricts eigenvalue estimates to the first P-2 eigenvalues. An excellent fit was obtained between predicted and observed values for all but the first eigenvalue by including the d_k term. Allen and Hubbard's (1986) empirical equations (hereafter referred to as the AH equations) have been implemented in the computer program PARALLEL (Hays, 1987).

Recently, Lautenschlager, Lance and Flaherty (in press) have improved upon Allen and Hubbard's general equation to provide a more nearly exact estimate of the first eigenvalue. They employed the following revised

equation:

$$[2]\quad \log(\hat{\lambda}_k) = \underline{a}_k + \underline{b}_k \log(\underline{N}-1) + \underline{c}_k \log\{(\underline{P}-\underline{k}-1)(\underline{P}-\underline{k}+2)/2\} + \underline{d}_k \log(\lambda_{k-1}) + \underline{e}_k \underline{P}/\underline{N}$$

which involves one additional predictor term in addition to the predictors used in equation 1, namely the ratio of the number of variables to the sample size. Allen and Hubbard (1986) reported an $R^2 = .931$ between the actual first eigenvalues and predicted first eigenvalues for their data using equation 1. Lautenschlager, et al. (in press) obtained an $R^2 = .933$ for their data using equation 1, and obtained an $R^2 = .993$ using equation 2. The increment in R^2 was statistically significant ($\underline{F}_{(1,\ 92)} = 782.01$, $\underline{p}<.0001$) using equation 2 over equation 1. Revised empirical regression equations (hereafter referred to as the LL equations) would appear to have the advantage of making better predictions of the first eigenvalue, and hence better predictions all around, owing to the fact that the regression estimates are recursive, meaning estimates of initial eigenvalues figure into estimates of later eigenvalues in applications of equations 1 and 2. Both equations produce regression estimates of eigenvalues that ostensibly could be used as criteria for a parallel analysis.

Available procedures for implementing PA criteria in practice were compared. This involved the examination of regression equation methods that can be used to estimate random data eigenvalues from known values of the sample size and number of variables. Problems inherent in the application of these regression estimation procedures are described, and a more accurate method for determining PA criteria in practice is presented.

Method

Procedure

Data were generated in a series of Monte Carlo simulations in which the number of variables (P) ranged from 5 to 50 in steps of 5, and sample sizes (N) were chosen as 50, 75, 100, 150, 200, 300, 400, 500, 750 and 1000, to reflect a range typically found in published research. For each N, P combination the SYSTAT (Wilkinson, 1986) analysis package was used to (a) generate random data for N cases on each of P variables from a N(0,1) population, (b) create a correlation matrix based on this data, and (c) conduct a principal components analysis of the correlation matrix. For values of $P > 10$, one hundred replications were generated; for values of $P \leq 10$, two hundred replications were generated. A total of 12,000 unique datasets were created.

Within each fixed N, P combination the results were combined over replications and averaged to produce an average first eigenvalue based on 100 values, an average second eigenvalue etc. These averaged values were then used to form tables of averaged eigenvalues. These datasets had been used by Lautenschlager, et al. (in press) under the restriction that only those N, P combinations which satisfied $N \geq 3P/2$ were involved in the development of the LL equations.

Proposed Comparisons

The generated data provided the empirical criteria for the comparison of the regression equation estimates using the AH and LL estimation equations. In addition, tabled values of eigenvalues were used to interpolate values for eigenvalues not specifically covered by the tables. The accuracy of these interpolations were investigated vis-à-vis the regression estimates by conducting additional Monte Carlo simulations to

serve as criteria. The root mean squared error (RMSE), defined as the square root of the average squared deviation of the interpolations and the regression estimates from the averaged eigenvalues obtained from these additional simulations, was calculated in each case to permit comparisons. Because the regression estimation methods only generate P-2 estimates for any given case, all RMSE calculations involved at most only the first P-2 eigenvalues.

Results

Problems with Regression Estimates of Eigenvalues

Those who wish to use AH or LL regression equation methods to estimate the eigenvalues of random data correlation matrices can use the empirical equations as presented in a table found in each of the respective papers. In order to produce AH regression estimates, one can use the PARALLEL program (Hays, 1987). The LL equations have been programmed and can be implemented through use of the PARANAL program (Lautenschlager, 1988). At first blush it would appear that the AH equations were at a distinct disadvantage in comparison with the LL equations, as the latter were derived from the generated data described above. Neither regression estimation method was found to be generally useful across the ranges of N and P from which the equations were developed.

Although for certain N, P combinations the results produced by regression estimation methods tended to agree with the Monte Carlo simulations, numerous other combinations produced clearly divergent, and often unreasonable results. As a case in point, fixing P = 50 (i.e. 50 variables were involved in the analysis) whenever N > 170 the AH equations begin to produce predictions of subsequent eigenvalues that were larger

than preceding eigenvalues (e.g. when N = 171 estimated eigenvalue #45 > estimated eigenvalue #44). As sample size grew larger while holding P = 50, even more peculiar things happened. Estimated eigenvalues decreased beyond the first value up to a point but then consistently increased. In addition, predicted eigenvalues were obtained for eigenvalues past the (P/2)th value that exceed the number of variables in the analysis. At the extreme high end where N = 1000, the estimated first eigenvalue was reasonably close to the empirical value. However, the estimate for the 48th eigenvalue was slightly over 3,849,433,795. Admittedly, this was at the extreme upper bound of both N and P, but similar problems occurred for other combinations of sample sizes and number of variables. Obviously the usefulness of these empirically obtained regression equations was not uniform across the N, P combinations studied by Allen and Hubbard (1986).

Were the LL equations any better at providing estimates of eigenvalues? The answer was a qualified yes. In general, the estimates based on the empirical equations provided by Lautenschlager, et al. (in press) were better behaved, but not well-behaved in terms of being generally useful to researchers. The LL empirical equations tended to produce better estimates than the AH equations, in the sense that estimates tended to be closer to the simulated empirical data (from which the estimation equations were developed), but these equations also sufferred from the problems described in the preceding paragraph. However, for some combinations of N and P the AH estimates were better. The differences in precision may have been due in part to the differences in the N, P combinations employed in the two studies and to the number of replications. Although the LL equations were technically "better" than the AH equations in terms of overall fit, they also posed problems for

researchers desiring to implement the PA criterion in the numerous research situations ostensibly covered by the range of N and P values used to derive the AH and LL equations. It should be noted that these pronounced problems described above, which were easily identified at the extremes, were also present elsewhere in a less obvious form.

Factors Affecting Regression Estimates of Eigenvalues

One might wonder why the regression estimates of eigenvalues can be so far off, even well within the bounds of N and P used to generate the equations. After all, the smallest R^2 reported by Allen and Hubbard accounted for over 93% of the variance of the first eigenvalue, and that reported by Lautenschlager, et al. accounted for over 99% of the variance. First, and obviously, the estimates provided by either set of empirically derived regression equations must produce estimates that are somewhat closer to the mean the set of specific eigenvalues of a given order (e.g. all first eigenvalues) used in generating the equation. Given the large R^2s, this may seem a trivial factor. However, what is estimated is the natural logarithm of the eigenvalue. Small differences in the estimation accuracy of a logarithm will have a more pronounced impact on the estimation accuracy of the eigenvalue itself.

Second, as noted by Lautenschlager, et al. (in press), errors in prediction are cumulative (and likely multiplicative) in effect. This is because a series of different, but dependent equations must be applied in sequence to estimate a range of eigenvalues. The prediction of the third eigenvalue must suffer since the prediction of first and second eigenvalues were both somewhat in error themselves. Because estimates of preceding eigenvalues must be used to estimate subsequent eigenvalues, the proportions of variance accounted for, as reported in the development of

the regression equations in these articles, are potentially deceiving regarding this deterioration.

A third source for lack of fit is likely due to the fact that eigenvalues must be estimated within a specific N, P combination. On the other hand, the empirical equations were developed by extracting an eigenvalue of a specific order from across all N, P combinations. Variance between N, P conditions for a particular order eigenvalue is clearly accounted for, but obviously some variance across eigenvalues within conditions is not captured in the process.

Another Option for Determining Parallel Analysis Criteria

Although the preceding discussion may seem to imply that the PA criterion can only be employed by those willing to conduct their own Monte Carlo simulations, there was another alternative. Given that current regression estimation methods could not generally be recommended across the range of the N, P values from which the equations themselves were developed, a better approach was desired. An alternative means for applying the PA method was derived through the direct use of the available simulated data themselves. Tables 1 through 10 present averaged eigenvalues from principal components analyses of the 12,000 random data correlation matrices. The tables reflect all N, P combinations described earlier, and were arranged so that each table covered a specific value of P and presented the average eigenvalues arranged in decreasing order, obtained over replications for each specific sample size. These tables represent an implementation of PA criteria that can be widely used.

Insert Tables 1 through 10

Relative Accuracy of Interpolations and Regression Estimates

It was decided that simple linear interpolation of non-tabled N, P combinations may prove sufficiently accurate for deriving PA criteria, and this hypothesis was examined in the sections below. The accuracy of the various estimation methods was tested by computing RMSEs for each method based on deviations from averaged eigenvalues of the same order obtained from Monte Carlo simulation results to cover each of the interpolation cases described below. Several specific examples and a number of general tests of the accuracy of the interpolation method were provided relative to the AH and LL regression equation estimates.

Examples of Linear Interpolation

The first interpolation case concerned a value for P which was covered by one of the tables, for a specific sample size that was not. For example, when P = 25 and N = 890 the last two columns in Table 5 were used. In this case the estimated first eigenvalue for random data was interpolated as:

$$1.345 + \{ ((750 - 890)/250) * (1.345 - 1.298) \} = 1.319$$

The AH estimated value in this case was 1.323 and the LL estimated value was 1.341. Subsequent eigenvalues were interpolated and regression estimates were obtained as well. These values could have been used to serve as PA criteria. A Monte Carlo simulation of 100 replications was conducted for this case (P = 25 and N = 890) and the results served as a baseline for computing RMSE values for the sets of eigenvalues estimated using linear interpolation, AH estimates and LL estimates. Since the regression methods could only estimate the first 23 eigenvalues, only the first 23 values were used in all RMSE calculations. The RMSE values obtained were

0.0019, 0.3069 and 0.1477, respectively. It should be noted in this case that both regression estimation methods produced 21 estimated eigenvalues greater than 1.0, and some of the latter AH estimates were greater than preceding estimates. The interpolation method was superior in this case.

A second interpolation case involved a value for P which was not covered by the tables, for a specific sample size that was covered. For example, when P = 17 and N = 75 the second columns of Tables 3 and 4 were used. In this case the estimated first eigenvalue for random data was calculated as:

$$2.050 + \{ ((17 - 20)/5) * (2.050 - 1.861) \} = 1.937$$

This value was closer to the LL regression estimated value of 1.914 than it was to the AH value of 2.003. Another Monte Carlo simulation involving 100 random samples for this particular case was used to develop expected eigenvalues. The RMSEs for the sets of eigenvalues obtained from linear interpolation, AH estimates and LL estimates were 0.0109, 0.0381 and 0.0264, respectively. Once again linear interpolation was more accurate, though the advantage was not as great as in the previous case.

A third interpolation case involved values for both N and P which were not covered by the tables. For example, when P = 37 and N = 177 the fourth and fifth columns of Tables 7 and 8 were used. In this case the estimated first eigenvalue for random data was calculated as:

Step 1:

$$2.058 + \{ ((150 - 177)/50) * (2.058 - 1.878) \} = 1.961$$

Step 2:

$$2.157 + \{ ((150 - 177)/50) * (2.157 - 1.976) \} = 2.059$$

Step 3:

$$2.059 + \{ ((37 - 40)/5) * (2.059 - 1.961) \} = 2.000$$

The last step produced the interpolated value. In this case the AH regression estimated value was 2.037 and the value was 1.938 for the LL method. Based on another Monte Carlo simulation of 100 replications for this case, the RMSEs for the sets of eigenvalues obtained by linear interpolation, AH estimates and LL estimates were 0.0055, 0.1682 and 0.2228, respectively. Again, linear interpolation proved more accurate than either of the regression estimation methods.

A More General Comparison of the Accuracy of Interpolations

Although the foregoing examples suggested that linear interpolation was relatively more accurate in a limited scope, it was important that the accuracy of this method be more definitively examined across a wide range of conditions. To this end the following additional comparisons were made. Since each table presents 10 separate conditions for a fixed value of P, it was possible to "interpolate" eigenvalues for the middle column of each group of three adjacent columns. For example, in Table 1 it was possible to interpolate values for N = 75 by using the columns for N = 50 and N = 100. The criterion for accuracy was then the respective column that has been interpolated. This resulted in a total of eighty sets of interpolations on values of N produced by interpolating for each of the middle eight columns in each of the ten tables. The benefit of interpolating for existing columns in the Tables was that it did not require any further Monte Carlo simulations. One drawback of this procedure was that the relative accuracy of the interpolation method was placed at a disadvantage, as interpolations were made over ranges of values for N much broader than would be done in practice. This type of interpolation is hereafter referred to as within-table interpolation.

Regression estimates for both the AH and LL equations were made for each of these conditions, and RMSEs were examined for the interpolations and regression estimates. In 71 out of the 80 cases the interpolated values produced the smallest RMSEs. A comparison of the relative sizes of the RMSEs for each method is presented in Table 11. Eighty-four percent of the interpolation RMSEs were less than or equal to .050, while only 34% or less of the regression estimates were as accurate. These results indicated that the interpolations were generally superior to either regression method.

For the 9 cases where a regression method performed better than the interpolation method, it was always the LL method. In all but two of these cases N < = 100, but there was no other discernable pattern to these nine cases. The largest discrepancy resulted from a case where the LL RMSE was .027 and the interpolation RMSE was .089. The average improvement in accuracy over these 9 cases was .019. The AH method never did better than linear interpolation. Focusing on the two regression methods only, the LL method was better in 54 cases.

Insert Table 11

It was also possible to conduct a series of interpolations for values of P across tables, for fixed values of N. For example, using Tables 1 and 3 it was possible to interpolate eigenvalues for P = 10 by using the same N column in each of the two tables. The criterion for accuracy was then the respective column in Table 2 that has been interpolated. Making all such interpolations of this kind produced eighty sets of interpolated

eigenvalues based on values for P in Tables 2 through 8. It should be noted that for these interpolations it was only possible to interpolate for eigenvalues up to the order of the smaller value of P used in the interpolation (e.g., when interpolating for P = 10 only 5 eigenvalues could be interpolated). This type of interpolation is hereafter referred to as across-tables interpolation.

Once again, regression estimates for both the AH and LL equations were made for each of these conditions, and RMSEs were examined for the interpolations and regression estimates. In 69 out of the 80 cases the interpolated values produced the smallest RMSEs. A comparison of the relative sizes of the RMSEs for each method is presented in Table 12. All 80 RMSEs for the interpolations were less than .050, while less than one third of the values for either of the regression methods were as accurate. Again it would appear that the interpolations were superior to either regression method.

For the 11 cases where a regression method performed better than interpolation, it was always the LL method. Eight of these eleven cases occurred for P = 10 involving all sample sizes except the two at either extreme. The largest discrepancy among those 11 cases resulted from an LL RMSE of .005 and an interpolation RMSE of 0.030. The average improvement in accuracy over all 11 such cases was .008, which indicated little actual difference between methods. The AH method never did better than linear interpolation. Focusing on the two regression methods only, the LL method was better than the AH method in 54 cases.

Insert Table 12

More Stringent Tests of Interpolation Accuracy

In practice one would rarely want to examine more than half the eigenvalues obtained, and more likely only the first third of the eigenvalues or less. So, as a further check on the accuracy of the interpolations vis-à-vis the regression estimation methods, an additional series of comparisons were made limiting the focus to only to the first third of the possible eigenvalues. These comparisons would tend to show the regression estimates in a much better light, however the linear interpolation method would still suffer the same disadvantage noted earlier in spanning at least twice the distance than would be required in actual use of the tables.

Under these circumstances, computing RMSEs for interpolations and regression estimates for only the first third of the possible eigenvalues, there was only a noticeable change for the within-table interpolations. Here the interpolation method was superior to either regression method in 53 out of the 80 possible cases. Of the twenty-seven cases where a regression method was better, 17 were for sample sizes of 75 or 100, with the remaining ten cases scattered from sample sizes of 150 to 500. For 21 of these 27 cases the LL regression method was best. Thus although the regression estimates improved when the proportion of eigenvalues estimated was reduced, and one of the regression methods proved more accurate for the smallest sample sizes, the interpolations were still generally better even for these stringent and somewhat biased tests.

It was interesting to note that for the across-tables interpolations there was virtually no improvement over the results obtained earlier when all estimated eigenvalues were involved. The interpolation method was superior for 69 of the 80 possible cases. For the 11 cases where a regression

method was better, eight of these cases involved P = 10 for the eight smallest sample sizes. The LL method was best in 10 of the 11 cases. This might have been anticipated in that the maximum number of eigenvalues that could be interpolated (and hence compared) was necessarily limited to the smaller value of P involved in the interpolation. In effect, the number of eigenvalues that were compared was already somewhat reduced in the initial set of comparisons.

It is important to point out that the within-table interpolations must necessarily estimate a large number of the exact same sets of eigenvalues as did the across-tables interpolations. In fact, sixty-four of the 96 possible interpolation cases were common to the two modes of interpolating values used here. Yet, a somewhat different picture of the accurracy of interpolations vis-à-vis the regression estimates occurred for the two modes of interpolation. Some of this difference in accuracy can be explained in terms of the range of values for P or N that were involved in a particular interpolation. Interpolations over values of P across-tables covered much narrower intervals, and were generally more precise than were interpolations involving values of N within a given table.

Comparisons Involving Adjusted Interpolation RMSEs

As noted previously, all the preceding comparisons, whether with all estimated eigenvalues or only the first third involved, had placed the interpolation method at a disadvantage. In practice one would interpolate between adjacent columns either within or across tables. Thus, in practice either type of interpolation would be much more accurate than in all the preceding comparisons! To attempt to control for this disadvantage a final series of comparisons were made. It was assumed

that the lack of fit for linear interpolations was due to two sources, random error and nonlinearity. Given the number of replications used to generate the tabled values, the first source was presumed a minor influence. In addition, it was assumed that even in the face of nonlinearity the linear interpolations would have to improve when adding a point that fell between two points, where all three such points were expected to fall on a specific monotonic curve. Thus, a considerable improvement in fit was expected since in practice there would be no need to skip over columns to make interpolations, as had been done above in the previous two sections (to avoid additional Monte Carlo simulations). It was further assumed that the linear interpolation RMSEs would potentially be reduced by a factor of either 25%, 33% or 50% of their previously estimated values. In view of a monotonic nonlinear relation it is quite likely that even more than a 50% improvement in RMSEs could be obtained. Though the exact amount of improvement expected was not known, these percentage reductions in RMSE values probably range from conservative to realistic.

Adjusted interpolation RMSEs were calculated based on the first third of the possible eigenvalues for the cases involving within-table interpolations. Under these conditions the linear interpolation method proved superior in 61 out of the 80 cases assuming only a 25% reduction in interpolation RMSEs could be expected. The 19 cases where one of the regression methods was more accurate included: values of $\underline{P}$ = 10 through 25, with sample sizes of 75 or 100; values of $\underline{P}$ = 30 through 40, for a sample size of 75; and values of $\underline{P}$ = 40 through 50, for a sample size of 100. The LL regression method was best in 14 of these cases. The number of cases where the linear interpolation method was superior increased to

66 and 69 assuming a 33% and a 50% reduction in interpolation RMSEs, respectively.

When similar adjustments were made to RMSEs for across-table interpolations the number of cases, out of 80 possible, where the interpolation method was superior increased to 75, 76 and 77 assuming a 25%, a 33% and a 50% reduction in interpolation RMSEs, respectively. In each case where a regression method was the best it was always the LL method.

Examining Redundant and Non-Redundant Interpolations

Because of the way in which interpolations were made for specific target columns in Tables 1 through 10, sixty-four of the 80 possible interpolations conducted within-tables and across-tables were redundant with one another. As such, it was useful to examine whether there were any cases where both types of interpolation for a specific combination of N and P were less accurate than the best regression method. When all eigenvalue estimates were used in the RMSE calculations this occurred in only 3 out of the 64 possible cases. Assuming only a 25% reduction in interpolation RMSEs led to 0 such cases out of the 64.

Still considering those cases where the two modes of interpolation were redundant with each other, and now limiting the RMSE calculations to only the first third of the eigenvalues led to 6 out of the 64 cases where a regression method was better than either mode of interpolation. Assuming a 25% reduction in interpolation RMSEs led to only 2 such cases, and assuming a 33% reduction led to 0 such cases. Of the 32 cases that were unique to one particular mode of interpolation, a regression method was best in only 5 of these cases. This number was reduced to 3 assuming a 25% reduction in interpolation RMSEs and to 1 assuming a 50% reduction.

Discussion

It seems reasonable to assume that the linear interpolation method for non-tabled values will generally be much better than either regression estimation method across the conditions covered by the tables. In all cases where either of the regression methods produced estimates of eigenvalues which were unreasonable (staying within the bounds of the N, P combinations used) the interpolation method produced reasonable and reasonably accurate estimates of eigenvalues. Even where regression estimates were reasonable, the interpolated values were more accurate in most cases. This was true in spite of the fact that many of the comparisons made did not make allowance for the fact that the interpolations spanned greater distances than would be needed in practice. One can be reasonably assured that linear interpolation of eigenvalues based on Tables 1 through 10 is comparable to conducting a Monte Carlo analysis to establish parallel analysis criteria. Tables 1 through 10 can be used to provide accurate random data eigenvalue estimates for use as parallel analysis criteria.

It is unlikely in practice that one would want to interpolate values beginning with the first eigenvalue, but rather beginning at some point where tabled values for a given eigenvalue appear close to the same order eigenvalue in the sample results at hand. This is because the parallel analysis criterion involves determining where the sample data eigenvalues become smaller than a random data eigenvalue of the same order. In effect, if one plots the eigenvalues from sample data along with the eigenvalues from random data based on the same N and P, the PA criterion for determining the number of components to retain is at (or before) the point where the two lines intersect. It seems clear from the above

results that use of the tabled values presented here should make implementation of the PA criterion more practical. Further Monte Carlo analyses within the ranges of N and P covered by these tables should be unnecessary. The interpolation method for determining parallel analysis criteria is generally more accurate than available regression estimation procedures, and relatively easy to apply within the boundaries of the tables presented. Though reliance on a single rule-of-thumb for determining the number of components to retain is ill-advised (cf. Cliff, 1988), those desiring to use the parallel analysis criterion in conjunction with other rules for determining the number of components to retain can now more readily do so.

References

Allen, S. J. & Hubbard, R. (1986). Regression equations for the latent roots of random data correlation matrices with unities on the diagonal. Multivariate Behavioral Research, 21, 393-398.

Bartlett, M. S. (1951). The effect of standardization on a χ^2 approximation in factor analysis. Biometrika, 38, 337-344.

Cliff, N. (1988). The eigenvalues-greater-than-one rule and the reliability of components. Psychological Bulletin, 103, 276-279.

Hays, R. D. (1987). PARALLEL: A program for performing parallel analysis. Applied Psychological Measurement, 11, 58.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. Psychometrika, 30, 179-185.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. Educational and Psychological Measurement, 20, 141-151.

Humphreys, L. G. & Montanelli, Jr., R. G. (1975). An investigation of the parallel analysis criterion for determining the number of common factors. Multivariate Behavioral Research, 10, 193-205.

Lautenschlager, G. J. (1988). PARANAL: A program for computing parallel analysis criteria for principal components analysis. Author.

Lautenschlager, G. J., Lance, C. E. & Flaherty, V. L. (in press). Parallel analysis criteria: revised regression equations for estimating the latent roots of random data correlation matrices. Educational and Psychological Measurement.

Lawley, D. N. (1965). Tests of significance for the latent roots of covariance and correlation matrices. Biometrika, 43, 128-136.

Montanelli, Jr., R. G. & Humphreys, L. G. (1976). Latent roots of random data correlation matrices with squared multiple correlations on the diagonal: A Monte Carlo study. Psychometrika, 41, 341-348.

Velicer, W. F. (1976). The relation between factor score estimates, image scores and principal component scores. Educational and Psychological Measurement, 36, 149-159.

Zwick, W. R. & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. Psychological Bulletin. 99, 432-442.

Table 1
Average Eigenvalues of Random N(0,1) Data for P = 5

	N =									
Root	50	75	100	150	200	300	400	500	750	1000
1	1.392	1.330	1.297	1.234	1.201	1.166	1.143	1.127	1.104	1.089
2	1.161	1.134	1.120	1.096	1.086	1.070	1.060	1.052	1.046	1.039
3	0.990	0.984	0.989	0.996	0.993	0.995	1.001	0.999	0.999	0.999
4	0.818	0.853	0.867	0.893	0.912	0.925	0.937	0.944	0.953	0.960
5	0.640	0.699	0.729	0.783	0.809	0.844	0.860	0.880	0.900	0.914

Table 2
Average Eigenvalues of Random N(0,1) Data for P = 10

	N =									
Root	50	75	100	150	200	300	400	500	750	1000
1	1.787	1.611	1.528	1.435	1.366	1.294	1.260	1.230	1.186	1.158
2	1.529	1.421	1.358	1.297	1.249	1.205	1.179	1.158	1.129	1.110
3	1.330	1.268	1.229	1.188	1.161	1.136	1.118	1.104	1.084	1.074
4	1.159	1.145	1.122	1.102	1.087	1.074	1.067	1.059	1.048	1.043
5	1.012	1.025	1.029	1.021	1.021	1.020	1.016	1.015	1.013	1.015
6	0.876	0.915	0.926	0.945	0.957	0.967	0.969	0.975	0.980	0.985
7	0.760	0.810	0.841	0.870	0.896	0.914	0.925	0.933	0.947	0.954
8	0.633	0.710	0.753	0.800	0.828	0.858	0.877	0.892	0.913	0.923
9	0.519	0.605	0.658	0.719	0.757	0.802	0.826	0.847	0.873	0.892
10	0.399	0.493	0.559	0.627	0.680	0.732	0.765	0.790	0.830	0.850

Table 3
Average Eigenvalues of Random N(0,1) Data for P = 15

	N =									
Root	50	75	100	150	200	300	400	500	750	1000
1	2.065	1.861	1.724	1.573	1.491	1.399	1.349	1.306	1.253	1.212
2	1.799	1.636	1.551	1.442	1.378	1.310	1.268	1.239	1.194	1.167
3	1.605	1.486	1.417	1.347	1.296	1.235	1.212	1.188	1.154	1.131
4	1.438	1.350	1.308	1.252	1.219	1.172	1.158	1.143	1.115	1.099
5	1.281	1.231	1.215	1.181	1.150	1.122	1.110	1.103	1.084	1.071
6	1.140	1.125	1.119	1.106	1.093	1.077	1.071	1.063	1.054	1.045
7	1.024	1.033	1.039	1.036	1.035	1.032	1.031	1.025	1.022	1.022
8	0.915	0.946	0.961	0.973	0.984	0.987	0.991	0.991	0.993	0.998
9	0.815	0.867	0.882	0.913	0.929	0.944	0.952	0.958	0.966	0.972
10	0.710	0.777	0.811	0.853	0.877	0.901	0.913	0.923	0.938	0.948
11	0.609	0.701	0.743	0.790	0.821	0.855	0.874	0.889	0.907	0.923
12	0.529	0.616	0.669	0.726	0.772	0.811	0.836	0.854	0.880	0.897
13	0.444	0.537	0.601	0.673	0.714	0.770	0.793	0.816	0.849	0.870
14	0.361	0.461	0.527	0.603	0.657	0.720	0.749	0.773	0.816	0.843
15	0.264	0.373	0.433	0.534	0.585	0.665	0.692	0.729	0.775	0.803

Table 4
Average Eigenvalues of Random N(0,1) Data for P = 20

	N =									
Root	50	75	100	150	200	300	400	500	750	1000
1	2.334	2.050	1.909	1.716	1.604	1.492	1.422	1.371	1.301	1.259
2	2.041	1.837	1.710	1.577	1.483	1.401	1.337	1.300	1.247	1.212
3	1.835	1.673	1.584	1.469	1.402	1.330	1.283	1.251	1.206	1.178
4	1.669	1.543	1.470	1.378	1.332	1.274	1.238	1.208	1.172	1.148
5	1.513	1.432	1.373	1.301	1.266	1.217	1.192	1.171	1.140	1.122
6	1.379	1.329	1.287	1.233	1.209	1.164	1.146	1.131	1.108	1.095
7	1.255	1.221	1.196	1.170	1.151	1.123	1.111	1.097	1.080	1.071
8	1.139	1.130	1.117	1.106	1.097	1.080	1.075	1.069	1.053	1.049
9	1.029	1.041	1.038	1.046	1.044	1.039	1.039	1.036	1.029	1.026
10	0.940	0.956	0.970	0.991	0.994	1.005	1.003	1.006	1.005	1.004
11	0.840	0.881	0.908	0.936	0.948	0.964	0.966	0.975	0.980	0.983
12	0.747	0.812	0.843	0.878	0.901	0.925	0.935	0.947	0.955	0.961
13	0.667	0.740	0.781	0.827	0.856	0.885	0.900	0.914	0.931	0.941
14	0.585	0.671	0.720	0.779	0.812	0.845	0.870	0.888	0.908	0.922
15	0.504	0.602	0.661	0.728	0.768	0.808	0.835	0.858	0.882	0.901
16	0.433	0.544	0.606	0.679	0.718	0.772	0.804	0.825	0.857	0.877
17	0.369	0.480	0.550	0.629	0.675	0.734	0.769	0.795	0.831	0.853
18	0.304	0.417	0.490	0.577	0.630	0.695	0.733	0.759	0.804	0.830
19	0.244	0.359	0.428	0.521	0.583	0.650	0.695	0.722	0.774	0.803
20	0.175	0.284	0.358	0.459	0.526	0.597	0.646	0.679	0.736	0.767

Table 5
Average Eigenvalues of Random N(0,1) Data for P = 25

	N =									
Root	50	75	100	150	200	300	400	500	750	1000
1	2.588	2.257	2.053	1.840	1.713	1.574	1.494	1.432	1.345	1.298
2	2.289	2.037	1.878	1.689	1.608	1.486	1.412	1.365	1.293	1.255
3	2.064	1.853	1.745	1.586	1.517	1.417	1.351	1.311	1.254	1.220
4	1.870	1.717	1.627	1.506	1.440	1.357	1.309	1.270	1.219	1.190
5	1.722	1.592	1.519	1.424	1.373	1.297	1.259	1.234	1.189	1.163
6	1.584	1.487	1.429	1.357	1.312	1.248	1.217	1.197	1.159	1.141
7	1.461	1.383	1.344	1.289	1.256	1.208	1.178	1.163	1.134	1.115
8	1.342	1.292	1.268	1.221	1.199	1.162	1.142	1.130	1.108	1.094
9	1.225	1.214	1.182	1.166	1.148	1.119	1.107	1.101	1.083	1.074
10	1.121	1.121	1.110	1.109	1.096	1.082	1.073	1.069	1.061	1.053
11	1.029	1.039	1.047	1.054	1.047	1.046	1.042	1.042	1.037	1.033
12	0.934	0.962	0.980	1.002	1.003	1.008	1.008	1.013	1.013	1.012
13	0.843	0.893	0.918	0.952	0.959	0.973	0.979	0.985	0.993	0.993
14	0.761	0.828	0.868	0.901	0.916	0.940	0.947	0.958	0.968	0.974
15	0.688	0.761	0.806	0.854	0.877	0.904	0.921	0.930	0.944	0.954
16	0.615	0.700	0.749	0.805	0.835	0.872	0.893	0.904	0.923	0.934
17	0.543	0.641	0.697	0.763	0.795	0.838	0.862	0.878	0.903	0.916
18	0.480	0.583	0.646	0.718	0.757	0.805	0.835	0.851	0.881	0.897
19	0.424	0.535	0.599	0.673	0.718	0.774	0.806	0.825	0.860	0.877
20	0.361	0.485	0.551	0.629	0.677	0.738	0.775	0.796	0.836	0.858
21	0.309	0.432	0.500	0.585	0.636	0.706	0.746	0.770	0.813	0.837
22	0.255	0.373	0.448	0.542	0.600	0.672	0.715	0.743	0.788	0.815
23	0.210	0.326	0.396	0.495	0.558	0.634	0.681	0.714	0.763	0.794
24	0.165	0.274	0.346	0.451	0.509	0.593	0.646	0.681	0.734	0.766
25	0.116	0.216	0.294	0.389	0.453	0.546	0.602	0.640	0.698	0.738

Table 6
Average Eigenvalues of Random N(0,1) Data for P = 30

Root	N = 50	75	100	150	200	300	400	500	750	1000
1	2.813	2.434	2.181	1.945	1.812	1.643	1.546	1.481	1.387	1.333
2	2.517	2.192	1.994	1.812	1.686	1.560	1.470	1.414	1.331	1.291
3	2.301	2.035	1.869	1.700	1.594	1.483	1.411	1.366	1.295	1.259
4	2.109	1.879	1.747	1.612	1.521	1.422	1.365	1.323	1.262	1.227
5	1.932	1.755	1.646	1.531	1.456	1.370	1.321	1.288	1.232	1.202
6	1.774	1.648	1.560	1.455	1.394	1.320	1.279	1.250	1.206	1.178
7	1.642	1.546	1.479	1.394	1.339	1.276	1.243	1.216	1.179	1.154
8	1.527	1.447	1.403	1.326	1.284	1.239	1.207	1.186	1.153	1.132
9	1.404	1.362	1.323	1.269	1.235	1.199	1.174	1.155	1.128	1.112
10	1.295	1.273	1.247	1.216	1.187	1.158	1.141	1.128	1.104	1.092
11	1.194	1.199	1.175	1.156	1.142	1.122	1.112	1.098	1.083	1.073
12	1.096	1.114	1.120	1.104	1.099	1.089	1.080	1.070	1.061	1.055
13	1.014	1.037	1.049	1.059	1.058	1.052	1.050	1.044	1.041	1.036
14	0.933	0.967	0.993	1.008	1.018	1.019	1.021	1.020	1.021	1.018
15	0.852	0.905	0.937	0.960	0.975	0.985	0.994	0.994	0.998	1.001
16	0.775	0.840	0.884	0.913	0.935	0.955	0.964	0.970	0.979	0.982
17	0.697	0.778	0.829	0.869	0.895	0.922	0.937	0.945	0.957	0.966
18	0.631	0.724	0.779	0.825	0.860	0.893	0.910	0.921	0.938	0.949
19	0.567	0.669	0.733	0.782	0.823	0.861	0.882	0.897	0.917	0.931
20	0.506	0.613	0.681	0.745	0.790	0.831	0.855	0.874	0.898	0.913
21	0.449	0.563	0.633	0.709	0.750	0.800	0.827	0.851	0.879	0.894
22	0.396	0.514	0.586	0.668	0.715	0.768	0.801	0.828	0.860	0.878
23	0.345	0.465	0.543	0.629	0.680	0.740	0.776	0.803	0.840	0.859
24	0.296	0.422	0.499	0.589	0.647	0.709	0.749	0.778	0.819	0.841
25	0.251	0.378	0.458	0.554	0.610	0.679	0.721	0.752	0.799	0.822
26	0.210	0.333	0.415	0.516	0.577	0.649	0.693	0.728	0.777	0.805
27	0.171	0.292	0.372	0.476	0.540	0.617	0.666	0.703	0.753	0.785
28	0.134	0.248	0.332	0.437	0.502	0.583	0.635	0.672	0.730	0.765
29	0.100	0.209	0.291	0.395	0.460	0.551	0.604	0.642	0.703	0.739
30	0.068	0.161	0.242	0.345	0.416	0.505	0.566	0.604	0.670	0.709

Table 7
Average Eigenvalues of Random N(0,1) Data for P= 35

Root	N = 50	75	100	150	200	300	400	500	750	1000
1	3.050	2.588	2.326	2.058	1.878	1.708	1.602	1.536	1.428	1.365
2	2.721	2.353	2.152	1.914	1.769	1.624	1.533	1.463	1.376	1.326
3	2.499	2.175	2.001	1.799	1.685	1.553	1.471	1.415	1.337	1.294
4	2.306	2.039	1.884	1.705	1.610	1.492	1.419	1.374	1.303	1.267
5	2.128	1.917	1.781	1.624	1.540	1.443	1.374	1.337	1.273	1.238
6	1.977	1.801	1.690	1.561	1.473	1.393	1.337	1.300	1.243	1.215
7	1.832	1.689	1.595	1.489	1.418	1.349	1.299	1.272	1.218	1.192
8	1.703	1.595	1.519	1.423	1.371	1.306	1.263	1.242	1.194	1.171
9	1.582	1.499	1.431	1.363	1.318	1.262	1.231	1.211	1.171	1.149
10	1.484	1.406	1.362	1.306	1.270	1.225	1.200	1.180	1.149	1.130
11	1.374	1.327	1.303	1.254	1.227	1.192	1.168	1.152	1.127	1.111
12	1.267	1.251	1.231	1.207	1.182	1.155	1.137	1.127	1.104	1.092
13	1.176	1.179	1.171	1.153	1.143	1.122	1.109	1.099	1.084	1.076
14	1.086	1.108	1.107	1.101	1.099	1.089	1.083	1.074	1.066	1.057
15	0.997	1.039	1.051	1.056	1.064	1.057	1.057	1.048	1.046	1.040
16	0.916	0.972	0.993	1.010	1.025	1.024	1.030	1.026	1.026	1.022
17	0.840	0.910	0.945	0.967	0.988	0.993	1.000	1.003	1.005	1.005
18	0.767	0.849	0.893	0.928	0.947	0.964	0.975	0.977	0.987	0.989
19	0.704	0.796	0.838	0.889	0.913	0.935	0.949	0.956	0.968	0.974
20	0.638	0.738	0.789	0.846	0.878	0.908	0.923	0.933	0.949	0.956
21	0.573	0.684	0.744	0.807	0.844	0.879	0.899	0.911	0.930	0.940
22	0.518	0.636	0.699	0.767	0.805	0.847	0.874	0.889	0.912	0.923
23	0.464	0.582	0.655	0.730	0.773	0.820	0.851	0.868	0.893	0.907
24	0.410	0.532	0.615	0.697	0.742	0.792	0.825	0.846	0.875	0.891
25	0.362	0.492	0.571	0.662	0.709	0.766	0.800	0.824	0.857	0.875
26	0.319	0.453	0.533	0.625	0.676	0.739	0.775	0.799	0.838	0.858
27	0.273	0.411	0.496	0.589	0.645	0.713	0.748	0.779	0.819	0.841
28	0.240	0.373	0.459	0.556	0.618	0.684	0.725	0.757	0.801	0.826
29	0.202	0.337	0.419	0.520	0.587	0.655	0.702	0.732	0.783	0.809
30	0.167	0.300	0.384	0.487	0.554	0.627	0.675	0.709	0.763	0.792
31	0.139	0.262	0.348	0.454	0.521	0.596	0.651	0.686	0.742	0.775
32	0.108	0.229	0.309	0.418	0.487	0.571	0.625	0.661	0.721	0.757
33	0.082	0.195	0.275	0.385	0.454	0.541	0.596	0.635	0.699	0.735
34	0.058	0.161	0.238	0.347	0.417	0.508	0.565	0.606	0.673	0.714
35	0.038	0.126	0.193	0.302	0.371	0.469	0.527	0.573	0.643	0.686

Table 8
Average Eigenvalues of Random N(0,1) Data for P = 40

	N =									
Root	50	75	100	150	200	300	400	500	750	1000
1	3.263	2.741	2.466	2.157	1.976	1.766	1.659	1.573	1.465	1.397
2	2.928	2.512	2.279	2.005	1.864	1.671	1.582	1.517	1.415	1.357
3	2.699	2.341	2.132	1.903	1.768	1.608	1.529	1.464	1.376	1.323
4	2.511	2.184	2.014	1.810	1.693	1.549	1.482	1.424	1.344	1.295
5	2.318	2.055	1.905	1.731	1.630	1.504	1.436	1.390	1.314	1.269
6	2.172	1.939	1.807	1.658	1.568	1.460	1.393	1.352	1.285	1.245
7	2.023	1.833	1.715	1.585	1.505	1.411	1.357	1.318	1.257	1.224
8	1.890	1.732	1.630	1.521	1.450	1.371	1.316	1.288	1.235	1.203
9	1.761	1.638	1.554	1.460	1.405	1.330	1.286	1.256	1.211	1.181
10	1.636	1.548	1.479	1.399	1.356	1.292	1.254	1.227	1.190	1.163
11	1.526	1.461	1.404	1.346	1.310	1.255	1.222	1.201	1.168	1.142
12	1.429	1.389	1.341	1.292	1.268	1.221	1.191	1.172	1.145	1.126
13	1.321	1.311	1.279	1.240	1.221	1.187	1.167	1.150	1.124	1.109
14	1.227	1.237	1.219	1.194	1.181	1.153	1.138	1.124	1.104	1.091
15	1.137	1.161	1.165	1.150	1.137	1.123	1.108	1.099	1.085	1.074
16	1.053	1.092	1.102	1.105	1.095	1.093	1.081	1.077	1.067	1.057
17	0.982	1.034	1.044	1.058	1.061	1.060	1.055	1.052	1.048	1.041
18	0.904	0.972	0.995	1.019	1.028	1.030	1.029	1.031	1.030	1.026
19	0.833	0.912	0.946	0.972	0.989	1.001	1.003	1.008	1.012	1.010
20	0.762	0.852	0.900	0.936	0.954	0.971	0.978	0.985	0.993	0.995
21	0.693	0.799	0.851	0.899	0.921	0.946	0.956	0.964	0.976	0.981
22	0.637	0.746	0.807	0.859	0.882	0.917	0.932	0.942	0.957	0.966
23	0.583	0.696	0.760	0.819	0.849	0.893	0.909	0.918	0.940	0.951
24	0.526	0.647	0.715	0.783	0.816	0.864	0.886	0.899	0.922	0.935
25	0.475	0.600	0.671	0.749	0.786	0.837	0.863	0.879	0.905	0.921
26	0.427	0.553	0.631	0.715	0.756	0.810	0.839	0.859	0.887	0.905
27	0.378	0.514	0.592	0.683	0.728	0.786	0.818	0.840	0.870	0.890
28	0.334	0.473	0.556	0.648	0.696	0.760	0.796	0.820	0.854	0.876
29	0.295	0.435	0.517	0.615	0.669	0.738	0.772	0.799	0.838	0.861
30	0.258	0.399	0.482	0.586	0.642	0.712	0.750	0.779	0.820	0.847
31	0.220	0.360	0.448	0.552	0.613	0.686	0.728	0.757	0.803	0.830
32	0.187	0.327	0.415	0.518	0.584	0.662	0.707	0.737	0.786	0.814
33	0.156	0.293	0.383	0.488	0.553	0.637	0.683	0.718	0.767	0.799
34	0.128	0.263	0.351	0.456	0.525	0.612	0.659	0.697	0.748	0.783
35	0.102	0.227	0.320	0.427	0.495	0.587	0.638	0.674	0.729	0.767
36	0.080	0.201	0.287	0.398	0.467	0.560	0.615	0.653	0.710	0.750
37	0.061	0.174	0.257	0.368	0.438	0.532	0.591	0.631	0.690	0.731
38	0.044	0.146	0.227	0.337	0.409	0.503	0.563	0.606	0.670	0.712
39	0.028	0.118	0.197	0.302	0.375	0.469	0.533	0.579	0.645	0.692
40	0.016	0.088	0.158	0.263	0.336	0.433	0.495	0.542	0.614	0.663

Table 9
Average Eigenvalues of Random N(0,1) Data for P = 45

	N =									
Root	50	75	100	150	200	300	400	500	750	1000
1	3.474	2.908	2.584	2.239	2.055	1.828	1.699	1.628	1.509	1.430
2	3.121	2.660	2.390	2.100	1.932	1.745	1.628	1.561	1.452	1.386
3	2.880	2.489	2.249	1.992	1.842	1.674	1.576	1.512	1.414	1.353
4	2.664	2.338	2.128	1.909	1.765	1.619	1.531	1.469	1.378	1.326
5	2.480	2.193	2.021	1.811	1.700	1.563	1.488	1.431	1.352	1.300
6	2.329	2.073	1.917	1.740	1.637	1.519	1.444	1.396	1.323	1.276
7	2.187	1.961	1.832	1.671	1.582	1.473	1.407	1.363	1.297	1.254
8	2.048	1.858	1.747	1.606	1.527	1.432	1.375	1.334	1.270	1.234
9	1.927	1.765	1.663	1.547	1.474	1.393	1.340	1.305	1.244	1.215
10	1.801	1.669	1.586	1.495	1.425	1.356	1.307	1.275	1.224	1.196
11	1.690	1.581	1.522	1.439	1.379	1.318	1.277	1.245	1.202	1.177
12	1.582	1.503	1.455	1.387	1.335	1.284	1.248	1.218	1.182	1.158
13	1.474	1.427	1.387	1.330	1.293	1.248	1.218	1.195	1.161	1.141
14	1.381	1.351	1.326	1.284	1.251	1.216	1.189	1.171	1.142	1.124
15	1.283	1.278	1.264	1.235	1.211	1.182	1.162	1.149	1.123	1.107
16	1.206	1.209	1.207	1.188	1.176	1.153	1.136	1.124	1.103	1.091
17	1.122	1.146	1.149	1.142	1.137	1.120	1.108	1.099	1.085	1.076
18	1.047	1.083	1.098	1.101	1.101	1.089	1.082	1.078	1.067	1.060
19	0.971	1.022	1.040	1.060	1.065	1.060	1.060	1.055	1.050	1.045
20	0.901	0.966	0.998	1.019	1.031	1.030	1.034	1.032	1.031	1.030
21	0.832	0.908	0.948	0.981	0.997	1.005	1.010	1.012	1.014	1.016
22	0.759	0.853	0.901	0.944	0.961	0.976	0.988	0.990	0.998	1.000
23	0.703	0.801	0.855	0.907	0.929	0.950	0.964	0.971	0.981	0.986
24	0.642	0.754	0.814	0.870	0.899	0.926	0.940	0.950	0.964	0.972
25	0.581	0.707	0.770	0.834	0.865	0.900	0.919	0.930	0.949	0.957
26	0.528	0.656	0.729	0.801	0.834	0.873	0.897	0.912	0.933	0.943
27	0.478	0.614	0.689	0.769	0.803	0.850	0.874	0.892	0.917	0.929
28	0.424	0.572	0.652	0.734	0.775	0.825	0.853	0.871	0.900	0.914
29	0.385	0.532	0.613	0.699	0.747	0.800	0.832	0.852	0.884	0.900
30	0.339	0.492	0.575	0.667	0.718	0.777	0.810	0.834	0.868	0.885
31	0.299	0.456	0.541	0.636	0.690	0.754	0.789	0.815	0.851	0.872
32	0.262	0.418	0.509	0.605	0.663	0.730	0.770	0.796	0.837	0.858
33	0.232	0.384	0.475	0.576	0.635	0.706	0.749	0.777	0.820	0.844
34	0.198	0.348	0.444	0.543	0.610	0.683	0.726	0.756	0.804	0.830
35	0.168	0.319	0.409	0.516	0.584	0.660	0.706	0.738	0.788	0.816
36	0.143	0.286	0.380	0.487	0.558	0.637	0.684	0.720	0.772	0.803
37	0.118	0.260	0.351	0.459	0.531	0.614	0.665	0.701	0.753	0.789
38	0.096	0.232	0.319	0.432	0.505	0.590	0.645	0.682	0.736	0.773
39	0.076	0.206	0.292	0.406	0.478	0.569	0.623	0.662	0.720	0.758
40	0.059	0.180	0.261	0.380	0.453	0.545	0.602	0.641	0.704	0.742
41	0.043	0.153	0.233	0.351	0.427	0.522	0.578	0.622	0.685	0.726
42	0.030	0.132	0.210	0.324	0.400	0.497	0.556	0.599	0.664	0.707
43	0.019	0.109	0.185	0.293	0.373	0.469	0.532	0.576	0.643	0.690
44	0.010	0.086	0.158	0.264	0.343	0.439	0.504	0.547	0.619	0.668
45	0.004	0.064	0.126	0.227	0.304	0.399	0.471	0.514	0.588	0.643

Table 10
Average Eigenvalues of Random N(0,1) Data for 50 Items

	N =									
Root	50	75	100	150	200	300	400	500	750	1000
1	3.675	3.044	2.712	2.331	2.124	1.885	1.759	1.678	1.533	1.457
2	3.336	2.804	2.500	2.188	2.015	1.802	1.685	1.612	1.483	1.415
3	3.073	2.621	2.372	2.079	1.928	1.730	1.628	1.558	1.445	1.382
4	2.861	2.474	2.247	1.985	1.849	1.675	1.579	1.511	1.412	1.352
5	2.665	2.334	2.137	1.905	1.776	1.623	1.531	1.476	1.382	1.327
6	2.493	2.216	2.036	1.829	1.714	1.573	1.492	1.442	1.355	1.305
7	2.355	2.092	1.941	1.755	1.655	1.526	1.455	1.408	1.330	1.283
8	2.212	1.997	1.852	1.690	1.607	1.482	1.420	1.373	1.305	1.262
9	2.083	1.888	1.771	1.627	1.552	1.445	1.385	1.345	1.281	1.241
10	1.968	1.794	1.695	1.570	1.501	1.406	1.353	1.316	1.257	1.222
11	1.844	1.703	1.622	1.520	1.455	1.370	1.324	1.291	1.237	1.205
12	1.737	1.621	1.556	1.465	1.409	1.339	1.295	1.263	1.215	1.187
13	1.639	1.546	1.490	1.416	1.368	1.304	1.268	1.238	1.197	1.171
14	1.528	1.465	1.428	1.368	1.322	1.273	1.236	1.216	1.178	1.154
15	1.433	1.394	1.365	1.319	1.279	1.239	1.209	1.191	1.157	1.139
16	1.347	1.322	1.304	1.270	1.241	1.209	1.185	1.168	1.139	1.122
17	1.252	1.263	1.249	1.226	1.204	1.178	1.160	1.146	1.121	1.107
18	1.169	1.197	1.196	1.185	1.165	1.150	1.134	1.125	1.103	1.091
19	1.081	1.134	1.146	1.139	1.130	1.121	1.110	1.101	1.087	1.077
20	1.006	1.074	1.096	1.099	1.090	1.091	1.086	1.079	1.068	1.062
21	0.938	1.015	1.045	1.060	1.057	1.067	1.063	1.060	1.051	1.048
22	0.873	0.962	0.998	1.020	1.025	1.039	1.041	1.039	1.037	1.033
23	0.797	0.907	0.948	0.984	0.997	1.012	1.018	1.017	1.021	1.020
24	0.740	0.852	0.899	0.949	0.966	0.988	0.994	0.997	1.005	1.006
25	0.680	0.803	0.854	0.912	0.936	0.961	0.971	0.977	0.988	0.993
26	0.622	0.757	0.813	0.877	0.903	0.935	0.949	0.957	0.972	0.979
27	0.571	0.711	0.774	0.843	0.874	0.912	0.928	0.938	0.956	0.966
28	0.521	0.668	0.736	0.808	0.846	0.887	0.910	0.918	0.941	0.952
29	0.474	0.623	0.697	0.776	0.815	0.863	0.886	0.900	0.926	0.937
30	0.432	0.578	0.660	0.744	0.787	0.839	0.867	0.881	0.909	0.923
31	0.385	0.539	0.620	0.714	0.761	0.816	0.845	0.863	0.894	0.910
32	0.343	0.501	0.584	0.683	0.732	0.793	0.826	0.844	0.879	0.897
33	0.307	0.465	0.552	0.655	0.707	0.770	0.806	0.827	0.864	0.883
34	0.273	0.433	0.519	0.624	0.681	0.747	0.785	0.809	0.848	0.869
35	0.237	0.398	0.490	0.596	0.654	0.725	0.765	0.791	0.834	0.856
36	0.207	0.365	0.458	0.567	0.631	0.705	0.746	0.773	0.819	0.843
37	0.176	0.337	0.430	0.539	0.606	0.683	0.726	0.756	0.804	0.831
38	0.149	0.307	0.401	0.513	0.582	0.660	0.706	0.738	0.790	0.817
39	0.125	0.278	0.374	0.486	0.557	0.639	0.687	0.721	0.774	0.804
40	0.106	0.248	0.345	0.463	0.531	0.617	0.667	0.702	0.757	0.792
41	0.085	0.223	0.320	0.437	0.508	0.594	0.648	0.684	0.742	0.778
42	0.066	0.198	0.292	0.412	0.483	0.572	0.628	0.666	0.726	0.763
43	0.050	0.175	0.269	0.388	0.457	0.551	0.608	0.647	0.709	0.749
44	0.036	0.153	0.244	0.361	0.435	0.529	0.588	0.628	0.694	0.734
45	0.024	0.132	0.219	0.334	0.410	0.507	0.568	0.609	0.678	0.720
46	0.015	0.114	0.195	0.311	0.387	0.485	0.547	0.589	0.660	0.704
47	0.008	0.095	0.174	0.286	0.364	0.462	0.523	0.570	0.642	0.687
48	0.003	0.077	0.149	0.262	0.337	0.437	0.500	0.549	0.622	0.670

Table 11
Ranges of RMSE Values for Interpolations on N Within Tables

Method of Eigenvalue Estimation	RMSE Range						
	≤.01	.01-.02	.02-.05	.05-.10	.10-.20	.20-.50	>.5
Linear Interpolation	34	20	13	12	1	0	0
LL Equations	4	8	15	19	17	10	7
AH Equations	0	1	16	16	15	17	15

Table 12
Ranges of RMSE Values for Interpolations on P Across Tables

Method of Eigenvalue Estimation	RMSE Range						
	≤.01	.01-.02	.02-.05	.05-.10	.10-.20	.20-.50	>.5
Linear Interpolation	47	23	10	0	0	0	0
LL Equations	2	10	14	17	21	8	8
All Equations	0	1	15	19	16	17	12

Author Notes

Correspondence concerning this article should be addressed to Gary J. Lautenschlager, Department of Psychology, University of Georgia, Athens, GA, USA 30602.